International Journal of Technological Exploration and Learning (IJTEL)
ISSN: 2319-2135
Volume 2 Issue 5 (October 2013)

Compress Compound Images in H.264/MPGE-4 AVC by Using Compound Image Coding

Girish Kumar Patra
Electronics and Communication Engineering
Qis College of Engineering and Technology
Ongole, Prakasam Dist, India

N Venkateswarlu
Electronics and Communication Engineering
Vizag Institute of Technology, Vizag

B Malleshwari
Electronics and communication Engineering
Qis College of Engineering and Technology
Ongole, Prakasam Dist, India

B Murali Krishna
Electronics and Communication Engineering
Vizag Institute of Technology, Vizag

Abstract— **Compound images are a combination of text, graphics and natural image. They present strong anisotropic features, especially on the text and graphics parts. These anisotropic features often render conventional compression inefficient. Thus, this paper proposes a novel coding scheme from the H.264 intraframe coding. In the scheme, two new intramodes are developed to better exploit spatial correlation in compound images. The first is the residual scalar quantization (RSQ) mode, where intrapredicted residues are directly quantized and coded without transform. The second is the base colors and index map (BCIM) mode that can be viewed as an adaptive color quantization. In this mode, an image block is represented by several representative colors, referred to as base colors, and an index map to compress. Every block selects its coding mode from two new modes and the previous intramodes in H.264 by rate-distortion optimization (RDO). Experimental results show that the proposed scheme improves the coding efficiency even more than 10 dB at most bit rates for compound images and keeps a comparable efficient performance to H.264 for natural images.**

Keywords- Base colors and the index map, compound image compression, dynamic programming, residual scalar quantization.

I. INTRODUCTION

Besides natural images, there are millions of artificial visual contents generated by computers every day, such as Web pages, PDF files, slides, online games, and captured screens. They are usually a combination of text, graphics, and natural image. This is why they are generally called as compound images. In particular, with cloud computing becoming more and more popular, compound images often need to be displayed on remote clients, wireless projectors, and thin clients. Some of the clients are unable to directly render them from files. Compressing and transmitting compound images provides a generic solution to these clients. However, in this solution, how to efficiently compress compound images has become a prevalent and critical problem.

The state-of-the-art image compression standards (e.g., JPEG, JPEG2000 and the intraframe coding of H.264) are all designed for natural images. The correlation among samples is mainly exploited by transforms (e.g., 2-D DCT or wavelet).Because of the complexity issue, 2-D transform is implemented by two 1-D transforms. Such separate implementation cannot handle anisotropic correlation rather than the horizontal and vertical.

Thus, approaches are proposed to efficiently exploit the anisotropic correlation among samples. One category of these approaches is to perform directional prediction before transform as H.264 intra frame coding does. After removing the directional correlation, prediction residues can be assumed as isotropic again. Another category is to apply directional transforms, such as directional wavelet and directional DCT which incorporate directional information into transforms, to reduce the high-frequency energy in the transform domain. They significantly improve the coding performance on natural images with rich anisotropic correlations. However, all of these approaches are not efficient enough when compressing compound images.

Taking the extreme anisotropic features of text and graphics into account, some schemes are proposed for compound image coding. The general ideas can be categorized into three groups. Image-coding-based approaches-They adopt conventional image coding schemes but improve the bit allocation between text/graphics and natural image areas because the text/graphics areas are often blurred after compression. Thus, the quantization steps in text/graphics areas are decreased and more bits are allocated to them. For a fixed bit budget, it would correspondingly decrease bits for the coding of natural image areas. Consequently, the overall quality after compression is still not good.

A. Layer-based Approach

They adopt the mixed raster content (MRC) image model for compression, where one compound image is decomposed into a foreground layer, a background layer, and a binary mask plane at block or image level. The mask plane indicates which layer each pixel belongs to and can be compressed by mature binary coding schemes, such as JBIG and JBIG2. The foreground and background layers are smoothed by data filling algorithms and then compressed by conventional image coding schemes. It demonstrates significant gains over conventional image coding schemes. However, there are several drawbacks in the approaches. First, the performance is greatly influenced by segmentation. Block-threshold and rate-distortion optimized methods are proposed to optimize the segmentation in [16] and [17]. Second, without special processing, the holes resulted from segmentation will deteriorate the coding performance. Third, separately coding text colors in the foreground layer and text shapes in the mask plane will also hurt the coding performance. A shape primitive extraction and coding (SPEC) scheme is proposed, where shape primitives containing both colors and shapes are losslessly compressed by a combined shape-based and palette-based coding algorithm.

B. Block-based approaches

They first classify blocks in compound images into different types according to their spatial properties [22]. Image features, such as histogram, gradient and the number of colors, are often used for classification, then different type blocks are compressed by Different coding schematic better adapt to their statistical properties. Considering the sparse histogram distribution of colors in text/graphics blocks, a novel method is proposed to represent a text/graphics block by several base colors and an index map. Furthermore, dingetal develop this method as an intra coding mode and incorporate it into H.264. thus, the coding performance of that scheme on compressing compound mages is significantly improved.

The proposed scheme in this paper adopts the block-based architecture as a basis. However, our focus is how to better exploit the spatial correlation in compound images without transform. H.264 intraframe coding is taken as the benchmark to develop our scheme, where the mode-based design and the rate-distortion optimization mode selection provide an easy way to combine spatial-domain and transform-domain methods. Our main contribution to develop a comprehensive and systematic coding scheme by fully taking the properties of compound images into account.

II. Proposed Compression Scheme

We would like to first analyze the properties of compound images before introducing our scheme. Different from natural images, compound images have their own characteristics especially on the text and graphics parts. To explain it clearly, Fig. 1 shows an exemplified 16 16 block with two letters "P" and "o". First, edges in compound images between letters and back-ground are much sharper than those in natural images. Some edges have several-pixel transition because of shadow effects, whereas the others do not have any transition. Second, geometries of edges are usually complicated and irregular. They are difficult to predict along a certain direction. Third, the block only has limited number of different sample values. For such text blocks, the intuitive feeling is that traditional transform will fail to give a compact representation in the transform domain. To verify it, we analyze the properties of text and graphics blocks in quantity by introducing two features: spectral activity measure (SAM) and spatial frequency measure (SFM).

Let us denote an image block as $x_{i,j,}$i=0,…,M-1 and j=0,…,N-1. M and N are the total numbers of samples in one column and one row, respectively. $F_{i,j}$ is the corresponding signal to $x_{i,j}$ in the frequency domain. SAM is a measurement of image predictability and it is defined in the frequency domain as

$$SAM = \frac{\frac{1}{M-N}\sum_{I=0}^{M-1}\sum_{j=0}^{N-1}\left|F(i,j)\right|^2}{\left\{\prod_{I=0}^{M-1}\prod_{J=0}^{N-1}\left|F(i,j)^2\right|\right\}^{\frac{1}{M-N}}} \qquad (1)$$

In this paper, we use DCT instead of DFT to get $F_{i,j}$ because it is the transform for compression in our scheme. SAM has a dynamic range of [1,∞]. Lower values of SAM imply lower predictability. SFM indicates the activity level of an image. It is defined as

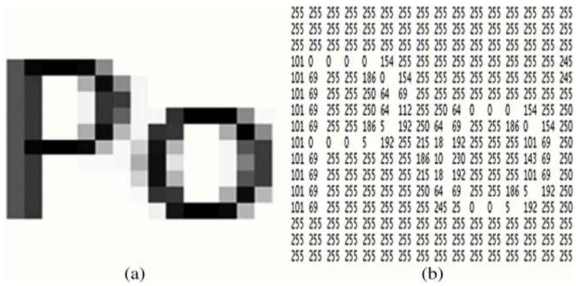

Figure 1. Exemplified 16 16 text block: (a) amplified block; (b) luminance samples values.

$$\begin{cases} SFM = \sqrt{R^2 + C^2}, \\ R = \sqrt{\frac{1}{M-N}\sum_{J=0}^{N-1}\sum_{i=0}^{M-1}(x_{i,j} - x_{i-1,j})^2} \\ C = \sqrt{\frac{1}{M-N}\sum_{i=0}^{M-1}\sum_{j=1}^{N-1}(x_{i,j} - x_{i,j-1})^2} \end{cases} \qquad (2)$$

R and C are defined as row and column frequency, respectively. They indicate the variations between two rows and between two columns. An image with high activity in the spatial domain will have a large value of SFM. Images with small SAM and large SFM are usually difficult to compress by transform because they are not predictable. With the two measurements, we analyze 3150 text/graphics blocks and 4552 natural image blocks of size 16×16. The SAM and SFM are calculated on each block. The histograms of SAM and SFM on those two types of blocks are depicted in Fig. 2. The SAM of the text and graphics blocks is concentrated in the low value areas and it indicates a low predictability of such blocks; the SAM of the natural image blocks is scattered over a much wider range. However, the SFM of the text/graphics blocks is scattered over high-value areas and it means high activities in the spatial domain. Meanwhile, the natural image blocks are much smoother and have a compact distribution of SFM in low-value areas. All these indicate that the text and graphics blocks are hard to compress efficiently by transform coding compared with natural images.

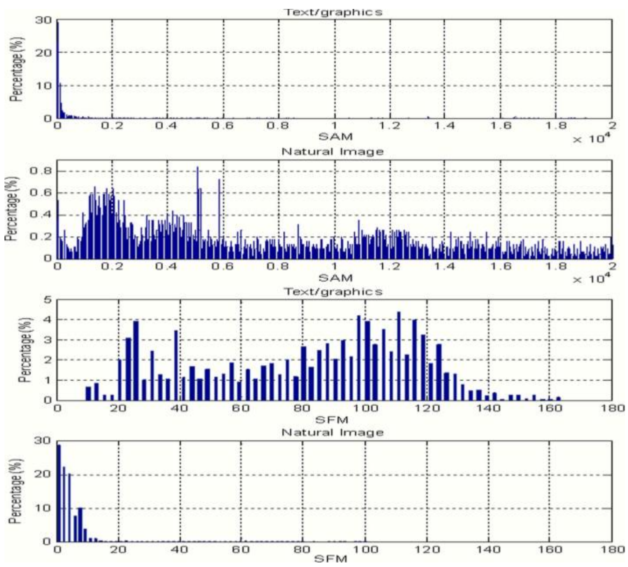

Figure 2. Histograms of SAM and SFM on text/graphics and natural image blocks.

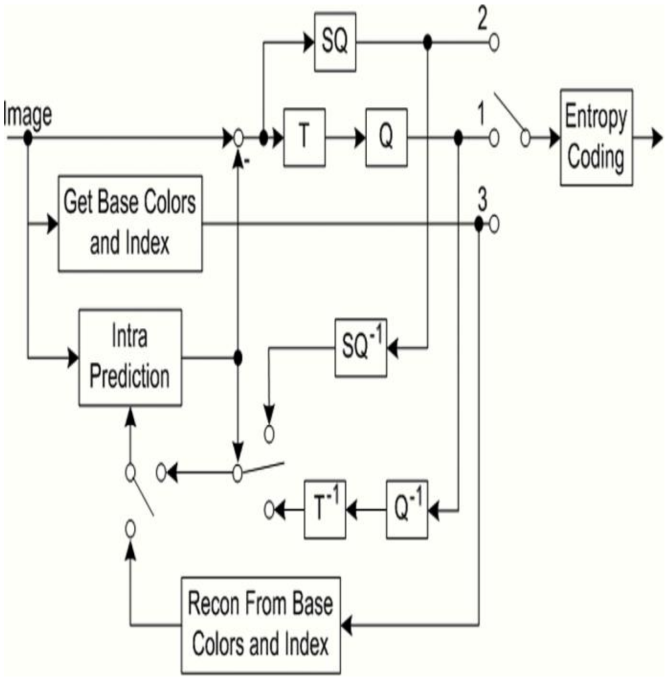

Figure 3. Block diagram of the proposed scheme.

If transform is removed from the compression of text and graphics block, what techniques can be used to exploit the spatial correlations among samples? One way is to directly compress the prediction residues by entropy coding without transform. It has been successfully adopted in H.264 intraframe coding and lossless intraframe coding. However, it is not applied to lossy intraframe coding of natural images yet because the residues still have some correlations that can be further exploited by transform. However, considering the extreme sharp edges and the surrounding shadows in text and graphics blocks as depicted in Fig.1,2,3, it should be more efficient than the transform coding. That motivates us to propose the spatial domain residual scalar quantization method to compress them.

A. *RSQ Mode*

For text and graphics blocks containing edges of many directions as shown in Fig. 1, intraprediction along a single direction cannot completely remove the directional correlation among samples. After intraprediction, residues still preserve strong anisotropic correlation. In this case, it is not efficient to perform a transform on them. One method is to skip the transform and directly code prediction residues, which is similar to traditional pulse-code modulation (PCM). However, the question is whether the performance of PCM is better than that of a transform for text and graphics residual blocks. To answer it, we introduce the method proposed to analyze the coding gain of PCM over a transform.

B. *BCIM Mode*

Having limited colors but complicated shapes is another property of the text and graphics parts on compound images. Such text/graphics blocks can be expressed concisely by several base colors together with an index map. It is somewhat like color quantization that is a process of choosing a representative set of colors to approximate all the colors of an image . In the BCIM mode, we first get the base colors of a block by using a clustering algorithm. All the base colors constitute a base color table. Then, each sample in the block will be quantized to its nearest base color. The index map indicates which base color is used by each sample.

Different from color quantization, each text/graphics block, but not an entire image, has its own base colors and an index map for representation in our scheme. Thus, it is content adaptive for each block. In addition, since the base color number of a block is small, fewer bits are required to represent each mapped index.

C. *Mode Selection and Mode Structure*

Each mode has its advantages at dealing with blocks of different features. One question that arises here is how to fully take advantage of each mode in the proposed scheme. It can be solved by the RDO algorithm that has been adopted by H.264. The best mode with the best block partition having the minimum rate-distortion cost will be selected to compress the current block.

All modes in the proposed scheme can be categorized into two types: spatial domain (SD) and DCT frequency domain (FD). There is a flag in the bit stream to distinguish them.. FD indicates the original intramodes in H.264, where the compression is performed in the DCT domain. SD indicates our proposed RSQ and BCIM modes. To adapt to the local nonstationary property of compound images, the spatial domain (SD) modes are applied to 16 × 16, 8× 8, and 4× 4 block sizes as those DCT frequency domain (FD) modes. The best mode in the spatial domain is compared with the best mode in the DCT frequency domain for the same size block in the rate-distortion sense. The better one is selected.

The DC mode in the spatial domain is replaced by the BCIM mode in the stream syntax. For those small size blocks, the BCIM mode is only performed on the luminance component in our scheme for simplicity. In 16× 16 blocks, the BCIM mode takes the place of the DC intramode. The better one between Dim3 and Dim1 is selected based on the rate-distortion criteria. For the Dim3 case, when the input image format is YUV 4:2:0, interpolation will be performed on UV color planes to get the same size color planes to facilitate the 3-D clustering. To be consistent with the block size of the DCT transform, the selection between direct quantization and transform coding on the 16 ×16 residual block, which is obtained by the prediction with three possible directions, is performed on 4 ×4 sub blocks.

III. EXPERIMENTAL RESULTS

We integrate the proposed methods into H.264/MPEG-4AVC reference software JM14.0 [37] to evaluate their performances.

Since the de-blocking filter in H.264 often blurs decoded compound images, it is disabled in our experiments. As shown in Fig., five captured screen images and a compound document are used as test images: fig 4 (a) is a web page, (b) is a snapshot of a typical screen scene, (c) is a slide, (d) is a combination of files, and (e) is a natural image. Their size is 1280× 1024. (f) is a compound document with size of 512× 768. Text and graphics are different in different compound images or in different regions of the same image. Some text and graphics blocks have no transition, whereas others have rich shadows. Some symbols on them are small and only occupy a 16× 16 or smaller size block but others may take up several blocks.

Experimental results of the PSNR on the luminance signal versus bit-rate curves are depicted in Fig. is set 47 to 7 with a step of 5. The results of H.264 integrated with only the proposed RSQ mode are marked by “RSQ” and those integrated with only the BCIM mode are marked by “BCIM”. The curves of H.264 with the RSQ and BCIM modes together are our proposed scheme, marked by “Proposed”. Another two schemes are selected for comparison: JPEG2000 and H.264 intraframe coding. For JPEG2000, we only compress the luminance plane by all bits with chrominance planes uncompressed. For the same bit budget, the results should be better than those all three color planes are compressed.

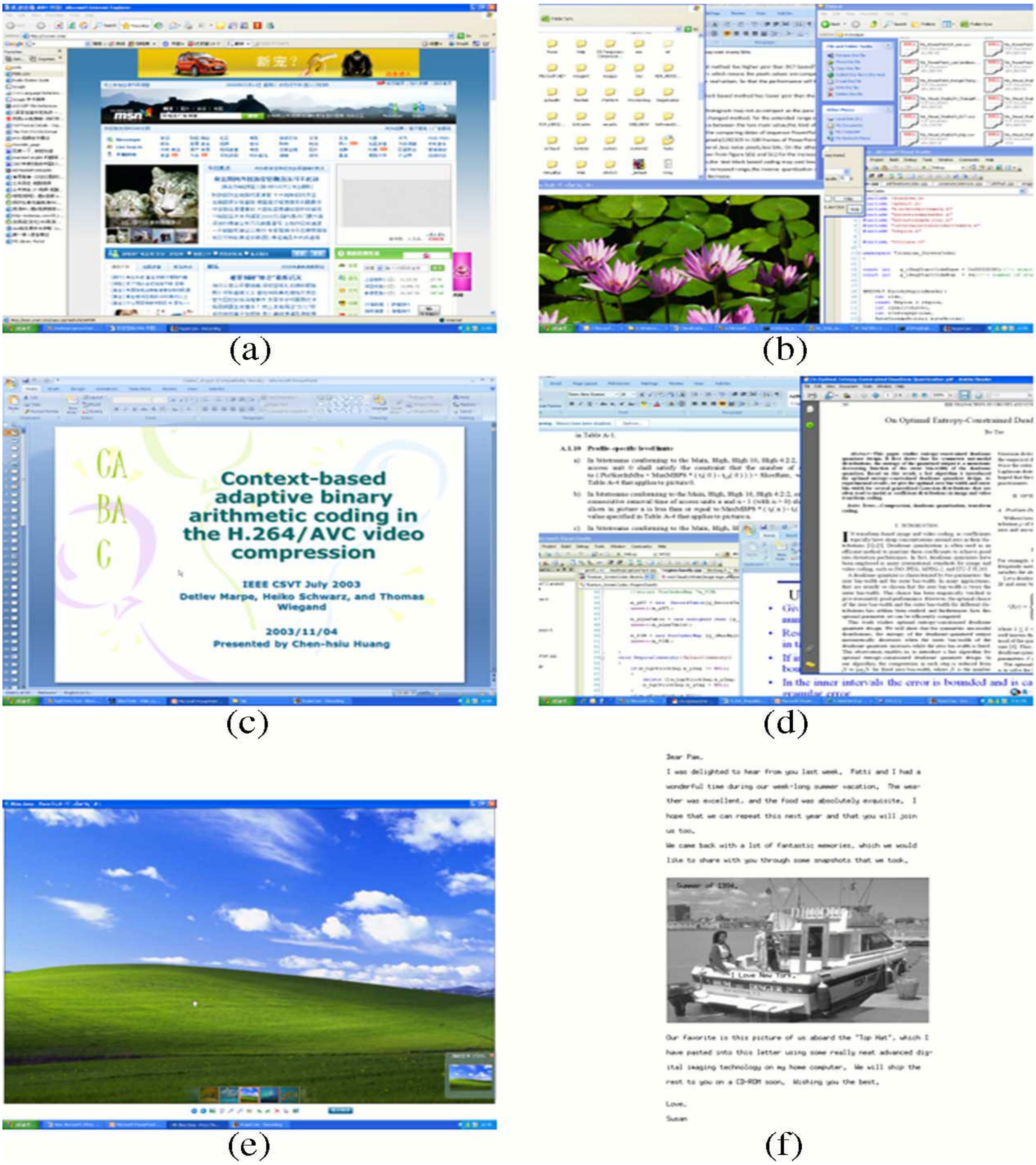

Figure 4. Six testing images/documents used in our (d) Files ; (e) Natural Image ; (f)Compound1

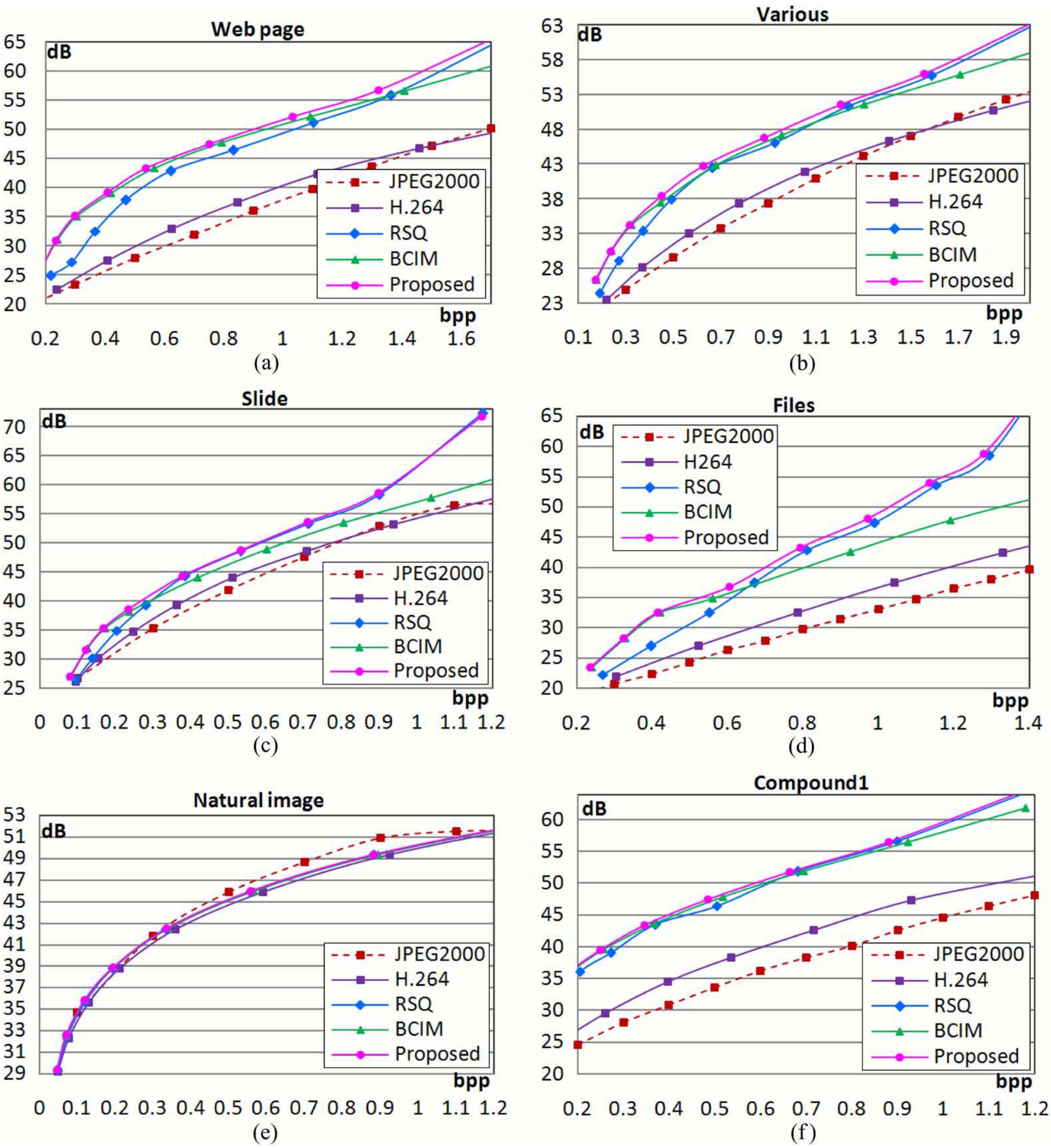

Figure 5. Experimental results on rate distortion curves.

As depicted in Fig., the BCIM mode has a better performance at low and middle bit rates and the RSQ mode has a better performance at middle and high bit rates for the first four compound images. By selecting the RSQ mode or the BCIM mode by RDO, the proposed scheme presents good performance at all bit rates. Compared with H.264 intraframe coding and JPEG2000, more than 10-dB gain at most bit rates is achieved in *Web Page*. Similar results are observed in the other compound images. For the fifth image, which is a natural image, the proposed scheme has a comparable performance to H.264. For *Compound1*, mode integrated.about 10 dB is obtained with RSQ mode or BCIM.

The percentages of the RSQ and BCIM modes in all testing images are given in Table I. They are obtained at three different rates: 0.4, 0.8, and 1.2 bpp. One can observe that the percentages of the RSQ mode increase with rate increasing, whereas the percentages of the BCIM mode decrease. But the phenomenon is not clear in *Natural Image*. It cannot be observed in *Compound1* because of many binary texts.

To evaluate the visual quality, parts of the magnified reconstructed images *Various* are shown in Fig. 9, at about 0.37 bpp. The images decoded by JPEG2000 and H.264 [(b) and (c), respectively] show severe blur and ring artifacts on the text and graphics parts. With the proposed two modes, the perceptual quality is greatly improved, close to the original image. The proposed scheme is also compared with Ding's approach [26]. Although the BCIM mode has the similar idea as that, the technology in the BCIM mode is improved significantly.

TABLE I. PERCENTAGE OF THE RSQ BCIIMM ODESAT DEFFERENT RATES

image		0.4bps	0.8bps	1.2bps
Web page	RSQ	3.3	7.2	10.2
	BCIM	28.4	21.9	17.6
Various	RSQ	4.0	8.7	9.8
	BCIM	29.5	17.4	16.9
Slides	RSQ	7.3	11.4	15.3
	BCIM	6.1	3.2	1.0
Files	RSQ	2.1	21.0	22.9
	BCIM	47.2	9.6	7.4
Natural image	RSQ	1.3	1.7	2.3
	BCIM	3.5	2.9	0.4
Compound *I*	RSQ	4.9	8.2	10.9
	BCIM	19.8	68.0	75.3

In Ding's approach, the rate cost is not considered and the base colors are directly selected by tree-structure vector quantization; it cannot achieve optimized rate-distortion performance. In BCIM mode, the rate cost is considered and the base colors are selected in the sense of RDO with clustering method of dynamic programming. Moreover, the BCIM mode is adaptive to different block sizes and component combinations. Experimental results of Ding's approach, the scheme with BCIM mode only and the proposed scheme are shown in Fig.5. It demonstrates the BCIM mode outperforms Ding's approach with a 1.4 Db gain at low and middle bit rates and more than 3 dB gain at high bit rates. Furthermore, the integration of the RSQ mode enables the proposed scheme to achieve a much higher performance at middle and high bit rates.

Finally, the complexity of the proposed scheme is discussed. Since the transform is skipped in the proposed RSQ and BCIM modes, the decoding complexity of the proposed modes is lower than that of the H.264 intramodes. At the encoder, the complexity of the RSQ mode is low too for the same reason. But the proposed BCIM mode needs to select base colors by clustering and its complexity is a little higher than that of H.264 intramodes. In the mode selection, the rate-distortion costs of all modes are calculated and then the mode with the minimum cost is selected for coding. Because of the proposed modes, the mode selection needs to check double the choices than it does in H.264 intracoding. It can, however, be significantly decreased by fast mode selection in future.

IV. CONCLUSION

We propose a compound image compression scheme by fully exploring spatial domain properties of compound images. Two spatial domain modes, called residual scalar quantization (RSQ) and base colors and the index map (BCIM), are integrated into H.264 intraframe coding; they achieve significant gains at all bit rates. The RSQ mode can cope with complicated text and graphics blocks in a simple way, which is just to quantize the intraprediction residues without a transform. The BCIM mode provides the ability to have a high performance improvement for the efficient representation form of the text/graphics block. They are both able to preserve the spatial structures of the text and graphics parts, important to visual quality. A rate distortion optimal method, similar to that in H.264, simplifies the mode selection and avoids the performance loss imported by the inaccurateness of segmentation. In short, this paper points out a good way to extend H.264 to compress compound images with simple technical extensions and to moderate complexity increasing because of addition mode selections.

REFERENCES

[1] Cloud Computing [Online]. Available: http://en.wikipedia.org/wiki/ Cloud computing

[2] W. P. Pennebaker and J. L. Mitchell, JPEG: Still Image CompressionStandard. New York: Van Nostrand Reinhold, 1993.

[3] D. Taubman and M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice. Norwell, MA: Kluwer, 2001.

[4] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC), , 2003, JVT of ISO/IEC MPEG and ITU-T, JVTG-050.

[5] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overviewbof the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003. [6] W. Ding, F. Wu, X. Wu, S. Li, and H. Li, "Adaptive directional lifting-based wavelet transform for image coding," IEEE Trans. Image Process., vol. 16, no. 2, pp. 416–427, Feb. 2007.

[6] C. L. Chang and B. Girod, "Direction-adaptive discrete wavelet transform for image compression," IEEE Trans. Image Process., vol. 16, no. 5, pp. 1289–1302, May 2007.

[7] H. Xu, J. Xu, and F. Wu, "Lifting-based directional DCT-like transform for image coding," IEEE Trans. Circuits Syst. Video Technol., pp. 1325–1335, Oct. 2007.